

World-wide Networking for LHC Data Processing

Artur Barczyk

California Institute of Technology, Pasadena, USA

Artur.Barczyk@cern.ch

Abstract: CERN's Large Hadron Collider is producing several Petabytes of physics data per year. We present the network environment used for LHC data processing, and provide outlook into evolution of computing models and networks supporting them.

OCIS codes: (000.0000) General; (000.0000) General

1. Introduction

The Large Hadron Collider at CERN [1] has started operation at 7 TeV beam energy in early 2010, since then delivering colliding proton beams to the four experiments (Atlas, CMS, Alice and LHCb). Data processing for the LHC is assured by the World-wide LHC Computing Grid (WLCG) [2] which provides a distributed computing platform aggregating the resources of nearly 150 centres around the globe. The computing centres are classified by function (and roughly by size of committed resources) as Tier-0 (CERN) producing the raw data from the detectors, 11 Tier-1 sites, and about 140 Tier-2 sites. In addition to the WLCG resources, about 300 Tier-3 sites, typically local computing clusters at the participating institutes are used in the physics analysis.

The LHC Optical Private Network (LHCOPN) is linking the Tier-0 (CERN) and Tier-1 sites, providing today an aggregate capacity of 140 Gbps in a partial mesh topology, as shown in Figure 1. The network is constructed as a set of Layer 2 point-to-point 10Gbps links interconnecting the data centers, with routing performed at the sites. Resiliency is provided either through multiple Tier-0 – Tier-1 links, or through use of Tier-1 – Tier-1 circuits providing backup (transit) connectivity in case of an outage on a primary link. The LHCOPN is open and neutral regarding the technologies used and implementation of the circuits between the sites. The implementations vary by site, e.g. NREN or GEANT lambda services are used in some instances, while the Tier-1s in US are connected to CERN through the dedicated US LHCNet network. In order to address the specific challenge of transatlantic connectivity, US LHCNet operates a SONET Layer 1 network, providing resilient Layer 2 services through mesh protection, with adjustable per service capacity using VCAT/LCAS SONET extensions.

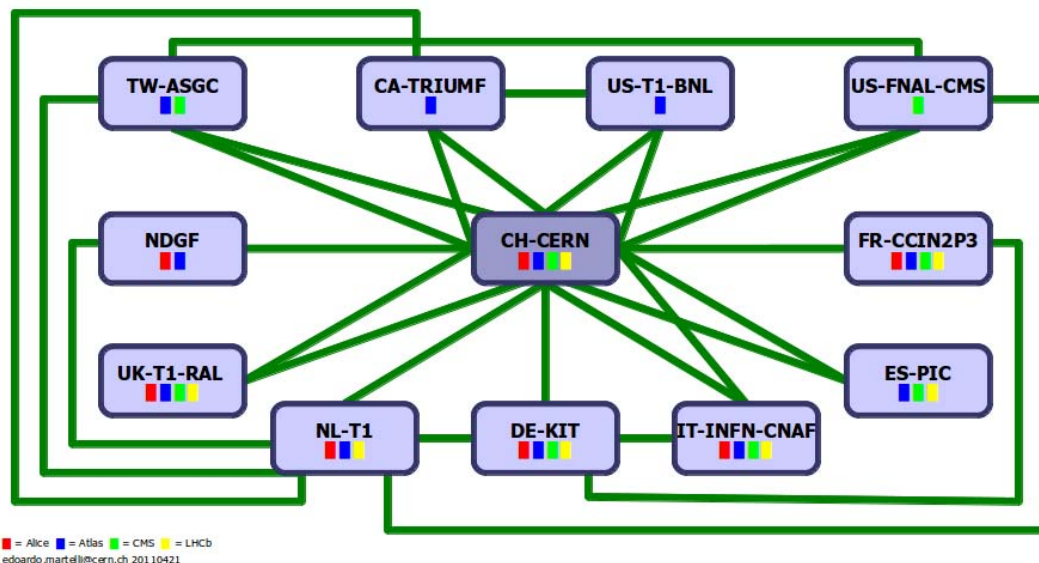


Figure 1: The current partial-mesh topology of the LHC Optical Private Network. Each light blue box indicates an LHC Tier-1 computing site. The dark box in the centre is the CERN Tier-0. Green lines indicate Layer 2 point-to-point circuits interconnecting the sites.

While the LHCOPN is dedicated to Tier-0 and Tier-1 data centres, most of the Tier-2 and Tier-3 sites use the general, routed IP connectivity, very often provided by their regional or national Research and Education Networks. Many of the larger Tier-2 sites have one or more 10Gbps uplinks today, and their number will grow in the near future.

2. Evolution of LHC Computing and Data Models

The original computing models have been relying on strategic data pre-placement and a strict tiered hierarchy of computing resources, as indicated on the left side of Figure 2. This approach has proved very successful, allowing very quick turn-around between data production and processing to data analysis performed by the physicists. It has also shown inefficiencies in particular in the storage resource allocation. The experience of first year of operation has also demonstrated the very high reliability of network resources available to the LHC community. This has led to developments in the computing models towards more dynamic data movement, driven more by the near real-time demand rather than pre-placement.

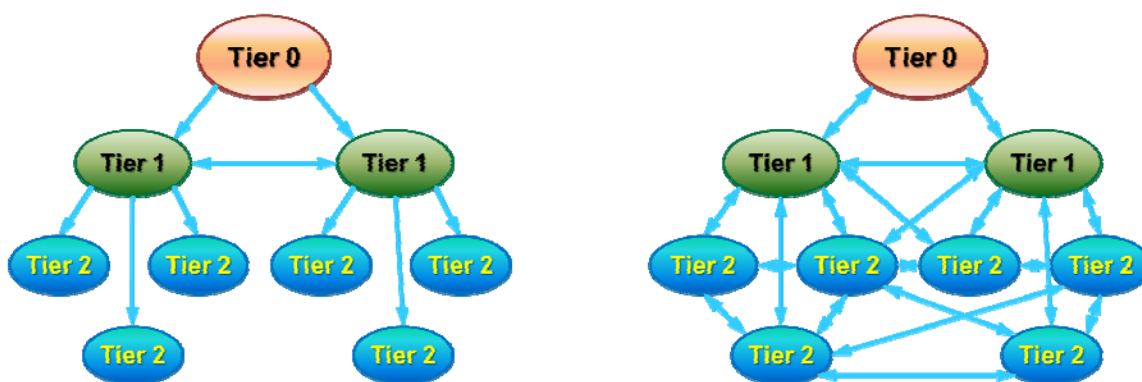


Figure 2: The strictly hierarchical MONARC model deployed at the beginning of LHC operation on the left, and the new model allowing for more dynamic, but less structured data movement and access.

The change in data models has an impact on the network requirements, as data movements becomes less predictable and many more sites, not connected today to the LHCOPN, require good network connectivity in order to fulfill their role.

The response to the new requirements is the LHC Open Network Environment (LHCONE) [3], a collaboration on global scale between R&E Networks and the LHC community. It is conceived as a multi-domain, hybrid overlay network, using resources provided by collaborating partners. LHCONE provides for a logical separation of LHC related traffic, allowing to Traffic-Engineer the networks with the goal to a) ensure the services to the LHC community maintain their quality and b) protect existing R&E infrastructures from very large data flows typically observed when LHC data is transferred between computing sites. LHCONE defines three services with different characteristics: a multi-point service, providing logical traffic separation; static point-to-point Layer 2 connections intended as high-capacity interconnects between major sites; and dynamic point-to-point Layer 2 connections, set-up for a user-specified duration and bandwidth allocation.

3. Hybrid Networks and Dynamic Bandwidth Allocation

The LHC experiments' computing models do not require full-mesh connectivity at full rate at all times, but for the overall system to work efficiently, data set transfer latency has to be kept at minimum, which requires high bandwidth at times. Dynamic bandwidth allocation therefore matches very well what's needed by the LHC data models.

Dynamic circuit services have been deployed in several R&E networks, e.g. ESnet has been operating a hybrid IP and dynamic circuit network since 2008 [4], providing circuit services to the LHC community among others. Internet2's ION and SURFnet's DRAC dynamic circuit services are two other examples of operational dynamic circuit networks where connected end-users can directly request network resources through an API or a Web GUI. One particular example of a system using dynamic circuits is DYNES [5], a virtual instrument being constructed in

the US, interconnecting up to 40 campuses through dynamic circuits. In Europe, GEANT's Bandwidth-On-Demand service AutoBAHN is expected to enter pilot phase in the next future, including the GEANT backbone as well as a number of selected European NRENs.

The Global Lambda Integrated Facility (GLIF) [6] has been promoting both static and dynamic lightpath technology and paradigm on a global scale. Many of the GLIF participants are actively involved in the definition of new standards relating to dynamic circuits, such as the OGF Network Services Interface (OGF-NSI) [7]. The GLIF Open Lightpath Exchanges (GOLEs) act as connection points between networks providing lightpath services, and in particular, more recently, are moving towards dynamically provisioned exchange-point connections tailored for multi-domain dynamic circuit services.

Taking into account the service availability in the most advanced R&E networks, as well as collaborative efforts in organizations such as GLIF, LHCONE will deploy dynamic circuit services intended to support scheduled or on-demand high-throughput data movement over guaranteed bandwidth channels between pairs of sites.

In order to take full advantage of the capabilities offered by dynamic circuit services, the LHC community needs to take a systems approach in integrating these new features in the experiments' computing models and software stacks. To cite two examples, the StorNet project [8], is developing an integrated end-to-end provisioning system, with co-scheduling of storage and network resources for efficient high-throughput data transfers; while the ESCPS project [9] addresses the "last mile" problem of dynamic circuit services at the campus and the integration with WAN circuit services.

4. Summary

Processing the LHC experiments' data is performed on a global infrastructure, with computing sites distributed on 6 continents. The computing and data models require robust, high-performance networks. This is achieved through collaboration between R&E networking and LHC communities, leading to the LHCOPN serving Tier-0 and Tier-1 sites, and the new LHC Open Network Environment (LHCONE), a hybrid overlay network to serve all LHC computing sites with both multipoint IP services as well as dynamically provisioned point-to-point lightpaths for high-throughput data movement.

5. References

- [1] CERN, <http://www.cern.ch>
- [2] WLCG, <http://lcg.web.cern.ch/lcg/public/>
- [3] LHCONE, <http://lhcone.net>
- [4] Inder Monga, Chin Guok, William E. Johnston, Brian Tierney, "Hybrid Networks: Lessons Learned and Future Challenges Based on ESnet 4 Experience", IEEE Comm. Mag., Vol. 49, No. 5, May 2011, pp 114-121
- [5] DYnamic NEtwork System, <http://www.internet2.edu/dynes>
- [6] GLIF, <http://www.glif.is>
- [7] Open Grid Forum – Network Services Interface (OGF-NSI), http://ogf.org/gf/group_info/view.php?group=nsi-wg
- [8] Shawn McKee et al., "StorNet: Integrated Dynamic Storage and Network Resource Provisioning and Management for Automated Data Transfers", to appear in proceedings of CHEP 2010, Journal of Physics: Conference Series (JPCS)
- [9] Phil DeMar, et al., "End Site Control Plane System Project", to appear in proceedings of CHEP 2010, Journal of Physics: Conference Series (JPCS)